Home Assignment 2

ECE 602 - Introduction to Optimization

Due: March 20, 2020

Exercise 1

(a) Find the conjugate functions for the following $f : \mathbb{R}^n \to \mathbb{R}$.

- $f(x) = I_C(x)$, where $I_C(x)$ is the indicator function of some non-empty, closed convex set C;
- $f(x) = \lambda \|\Sigma x\|_1$, where $\lambda > 0$ and $\Sigma \in \mathbb{S}^n_{++}$ is a diagonal matrix with its diagonal elements equal to $\sigma_i > 0$;
- $f(x) = \lambda \|\Sigma x\|_2$, where $\lambda > 0$ and $\Sigma \in \mathbb{S}^n_{++}$ are as above;
- $f(x) = \frac{1}{2} x^T A x + b^T x + c$, where $A \in \mathbb{S}^n_{++}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{S}$.
- $f(x) = -\sum_{i=1}^n \log x_i$, with $\operatorname{dom} f = \mathbb{R}_{++}$.

(b) Provide closed-form expressions for the proximal operators of all the above functions. (Use Moreau decomposition where needed.)

Exercise 2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function of $x \in \mathbb{R}^n$, and let its associated *proximal operator* be $\operatorname{prox}_f(x)$ (which is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$).

(a) Assuming that $f(x)$ is separable, *viz.*

$$f(x) = \sum_{i=1}^n f_i(x_i),$$

derive an expression for $\operatorname{prox}_f(x)$ in terms of $\operatorname{prox}_{f_i}(x_i)$.

(b) Let $W \in \mathbb{R}^{n \times n}$ be an orthogonal matrix (i.e., $W^T W = W W^T = I$). Derive an expression for the proximal operator of $f(Wx)$ in terms of $\operatorname{prox}_f(x)$. How does the expression changes when f is separable?

(c) Let $y \in \mathbb{R}^n$ be a measured signal contaminated by additive noises. For a properly chosen (square and orthogonal) W, the problem of *de-noising* of y can be formulated as an optimization problem of the form

$$x^* = \arg\min_x \left\{ \frac{1}{2}\|y - x\|_2^2 + \lambda \|Wx\|_1 \right\},$$

where $\lambda > 0$ is a user-defined regularization parameter. Derive a closed-form solution to the above problem using Moreau decomposition.

Exercise 3

Let I_0 be a grey-scale image represented by a real-valued matrix of size $n \times n$. Suppose that only half of the values of I_0 are known, while the other half has to be recovered through an optimization procedure. In what follows, you will be asked to compare the results produced by two alternative approaches.

Let Ω denotes the set of pixel coordinates over which the values of I_0 are known. Then, one can estimate the remaining values of the image by solving the following optimization problem

$$\begin{aligned} &\underset{I}{\text{minimize}}\ f_0(I) \\ &\text{subject to } I(i,j) = I_0(i,j),\ (i,j) \in \Omega. \end{aligned}$$

Note that, in this formulation, f_0 could be viewed as a "roughness measure" representing our a priori belief as to how variable the final result should be.

In this exercise, we consider two possible choices of f_0 as given below.

1. *Tikhonov regularizer (TR)*

$$f_0(I) = \sum_{i=2}^{n} \sum_{j=2}^{n} |I(i,j) - I(i-1,j)|^2 + |I(i,j) - I(i,j-1)|^2.$$

2. *Anisotropic total variation (aTV)*

$$f_0(I) = \sum_{i=2}^{n} \sum_{j=2}^{n} |I(i,j) - I(i-1,j)| + |I(i,j) - I(i,j-1)|.$$

Compute the above two solutions using the `cvx` package (`http://cvxr.com/cvx/`). Which of the two approaches provides a more natural result?

As an input, use a 64×64 image `I0` defined as

```
f = imread('cameraman.tif');
f = double(f(33:96,81:144));
```

Also, define Ω by means of a binary mask `M` given by

```
rng(2000);
M = false(64);
ind = randperm(64*64);
M(ind(1:64*64/2)) = true;
```

In this case, the observed values of I_0 are simply given by `I0(M)`, which is a column vector of length 2,048.

Finally, to visualize images, use

```
imagesc(I0)
axis square
colormap gray
```